Hippocampal shape analysis in Alzheimer's disease using Functional Data Analysis

Irene Epifanio[a]*, Noelia Ventura-Campos[b]

The hippocampus is one of the first affected regions in Alzheimer's disease. Left hippocampi of controls, mild cognitive impairment and Alzheimer's disease patients are represented by spherical harmonics. Functional data analysis is used in the hippocampal shape analysis. Functional principal component analysis and funcional independent component analysis are defined for multivariate functions with two arguments. A functional linear discriminant function is also defined. Comparisons with other approaches are carried out. Our functional approach gives promising results, especially in shape classification. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: Functional data analysis; Shape analysis; Alzheimer's disease; Principal component analysis; Independent component analysis; Discriminant analysis

1. Introduction

The early diagnosis of Alzheimer's disease (AD) is a crucial issue in our society, because the administration of medicines to individuals who are subtly impaired may render the treatments more effective. Mild cognitive impairment (MCI) is considered as a diagnostic entity within the continuum of cognitive decline towards AD in old age [1, 2]. Longitudinal studies show a direct relation between the hippocampal volume decrease and cognitive decline [3, 4]. However, volumetric measurements are simplistic features and structural changes at specific locations cannot be reflected in them. If morphological changes could be established, then this should enable researchers to gain an increased understanding of the condition. This is the reason why nowadays shape analysis is of a great importance in neuroimaging [5].

Several shape modeling approaches have been considered in the neuroimaging literature. One of them is the medial representation, where the binary object is represented using a set of atoms and links that connect the atoms together to form a skeletal representation of the object. Styner et al. [6] applied this scheme to hippocampi and other human brain structures. The distance map approach has been applied in classifying a collection of hippocampi in Golland et al. [7]. In a distance map, the distance from each point in the image to the boundary of the object is computed. Other approaches include deformation fields obtained by warping individual substructures to a template, such as the paper of Joshi et al. [8] applied to the hippocampus, or the landmark approach used with hippocampi by Park et al. [9] or Shen et al. [10]. Instead of the previous non-parametric models, a parametric approach, which has been successfully applied to model various subcortical structures, is the spherical harmonic representation (SPHARM) [11, 12, 13, 14, 15, 16, 17]. Each individual surface is parameterized by a set of coefficients weighting the basis functions: the spherical harmonics or its weighted version (the weighted spherical harmonic representation). Styner et al. [5] compared the sampled boundary implied by the SPHARM description with the medial shape description, obtaining good concordance between both descriptions. Other works propose global features which discriminate the condition [18]. It is quite common to use some of these approaches with a principal component analysis for shape classification and group comparison.

The spherical harmonic representation is a particular case of representing functional data as smooth functions. The whole surface is modeled from a set of points belonging to the surface. Functional data are observed discretely although a continuous function lies behind these data. In order to convert the discrete observations into a true functional form, each

[a] *Dept. Matemàtiques, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló, Spain*
[b] *Dept. Psicologia Bàsica, Clínica i Psicobiologia, Universitat Jaume I, Spain*
* *Correspondence to: Tel.: +34-964728390, fax: +34-964728429. E-mail: epifanio@uji.es*

function is approximated (smoothed) by a weighted sum (a linear combination) of known basis functions. Functional data analysis (FDA) provides statistical procedures for functional observations (a whole function is a datum). The goals of FDA are basically the same as those of any other branch of statistics. Ramsay and Silverman [19] give an excellent overview. Ferraty and Vieu [20] provide a complementary and very interesting view on nonparametric methods for functional data. A mixture of practical and theoretical aspects is found in Ferraty and Romain [21]. The field of FDA is quite new and there is still a lot of work to be done, but in recent years several applications have been developed in different fields, especially in human health [22, 23, 24].

In Epifanio and Ventura-Campos [25] two-dimensional (2D) shapes were analyzed from the three point of views considered by Stoyan and Stoyan [26] for describing shapes: firstly, set descriptors; secondly, using landmarks (point description); and thirdly, employing a function describing the contours. The results were compared with these three approaches (the set theory approach, the landmark based approach and the functional approach) in two of the main problems in form statistics: the study of the main sources of variation among the shapes (principal component analysis, PCA), and classification among different classes (discriminant analysis). The analysis of contour functions by FDA gave more meaningful results in both problems.

In this work, the hippocampus surface is described by multivariate (three) functions with two arguments. In Section 2, the methodology is introduced together with our data. We discuss the extension of the PCA to deal with trivariate functional data with two arguments. A discriminant function based on independent component analysis (ICA) is defined for indicating where the differences between groups are and what their level of discrimination is. In Section 3, the methodology is applied to the analysis of structural magnetic resonance imaging (sMRI) scans for studying the hippocampal differences among the subjects of three groups: cognitively normal (CN) subjects, patients with mild cognitive impairment (MCI), and patients with early Alzheimer's disease (AD). Comparison with other works is carried out. Finally, conclusions and some open problems are discussed in Section 4.

2. Materials and methods

2.1. Brain sMRI scans processing

A total of 28 individuals (12 CN, 6 MCI and 10 AD subjects) are analyzed in this study, whose description is in Table 1. All the individuals were recruited from the Neurology Service at La Magdalena Hospital (Castelló, Spain) and the Neuropsychology Service at the Universitat Jaume I. All experimental procedures complied with the guidelines of the ethical research committee at the Universitat Jaume I. Written informed consent was obtained from every subject or their appropriate proxy prior to participation. Selection for the participant group was made after careful neurological and neuropsychological assessment. The neuropsychological test battery involved Digit Span, Similarities, Vocabulary, and Block Design of the WAIS-III; Luria's Watches test, and Poppelreuters Overlapping Figure test. sMRI were acquired on a 1.5T scanner (General Electric). A whole brain high resolution T1-weighted anatomical reference scan was acquired (TE 4.2 ms, TR 11.3 ms, FOV 24 cm; matrix = $256{\times}256{\times}124$, 1.4 mm-thick coronal images).

Hippocampi were traced manually on contiguous coronal slices (or sections) following the guidelines of Watson et al. [27], and Hasboun et al. [28]. The hippocampus segmentation was done by an expert tracer with MRicro software, blinded to the clinical data of the study subjects. The segmentation of each hippocampus lasted approximately 40 minutes. An example of the left and right hippocampal contour (drawn in white) in a coronal view is shown in Figure 1 (a), while a sagittal view of one of the hippocampus can be seen in Figure 1 (b). Images were visually reoriented. The anterior commissure–posterior commissure (ACPC) line was identified and the images were then reoriented parallel to it. The slices were put together using the isosurface function in Matlab, which gives the vertices and faces of the triangle mesh.

2.2. Surface parametrization

In 2D, the contour is a closed planar curve that consists of the elements of the figure boundary. The contour parametrization by its arc length can be applied to any contour (note that other contour functions have limitations, see Kindratenko [29] for a review of various contour functions).

This method has been extended to represent analogously the surfaces of closed 3D objects. In this case, instead of two parametric functions with one parameter, three functions with two angular parameters are needed: $x(\theta, \phi)$, $y(\theta, \phi)$, $z(\theta, \phi)$ (see L. Shen and McPeek [15] for a detailed explanation). Specifically, a surface is mapped onto a unit sphere under a bijective mapping. However, unlike the 2D case, some practical problems prevent this mapping from being completely straightforward. In fact, this one-to-one mapping can be obtained from various surface flattening techniques such as conformal mapping [12], semi-isometric mapping [30], area preserving mapping [13, 31] and the deformable surface algorithm [32]. Since the conformal mapping tends to introduce huge area distortion, area preserving mapping is widely used. However, these flattening methods are computationally intensive and not trivial to implement. Here, we use a new

alternative proposed recently in Chung et al. [17] for objects that are close to either star-shape or convex. The mapping is based on the equilibrium state of heat diffusion. The idea is tracing the geodesic path of heat equilibrium state from a heat source (hippocampus in this case) to a heat sink (sphere). As solving an isotropic heat equation in a 3D image volume is computationally trivial, this flattening technique is numerically simpler than any other available methods and does not require optimizing a cost function. Details about this method can be found in Chung et al. [17], and how it works can be seen at http://www.stat.wisc.edu/~mchung/research/amygdala/. This step although necessary is auxiliar, and any other method can be used without changing the following analysis. However, the method from Chung et al. [17] has been effective with hippocampi, and we will use it.

Once the surface is mapped onto the sphere, the angles serve as coordinates for representing hippocampus surfaces using basis functions. Figure 2 shows an illustration of the surface flattening process for one hippocampus with that method, and the surface parameterization using the angles (θ, φ). The point $\theta = 0$ corresponds to the north pole of a unit sphere.

2.3. Representing functions by basis functions

The first step in FDA is the conversion from discrete data to functions by smoothing. Linear combinations of basis functions are used for representing functions. We have chosen as basis the spherical harmonics, because they have been already used in similar structures with excellent results, furthermore its orthogonality has computational advantages. Other possible basis could be the weighted Fourier series [14], spherical splines [33, 34] or spherical wavelets [35, 36], although spherical harmonics is the most used basis in this field.

Although complex-valued spherical harmonics could be used as in Gerig et al. [11] or Shen et al. [13], we have preferred to used real spherical harmonics as in Chung et al. [14, 17], considering that most applications of spherical harmonics require only real-valued spherical functions, and for convenience in setting up a real-valued stochastic model.

A real basis of spherical harmonics is given by (l is the degree and m is the order):

$$Y_{lm}(\theta, \varphi) = \begin{cases} \sqrt{2} N_{(l,m)} cos(m\varphi) P_l^m(cos\theta) & \text{if } m > 0 \\ N_{(l,0)} P_l^0(cos\theta) & \text{if } m = 0 \\ \sqrt{2} N_{(l,|m|)} sin(|m|\varphi) P_l^{|m|}(cos\theta) & \text{if } m < 0 \end{cases} \qquad (1)$$

where $N_{(l,m)} = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}}$ and P_l^m is the associated Legendre polynomial of order m defined over the range $[-1, 1]$: $P_l^m(x) = \frac{(-1)^m}{2^l l!}(1 - x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}}(x^2 - 1)^l$.

Let S^2 be the unit sphere in $\mathbb{R}^3$, and f and $g \in L^2(S^2)$. The inner product is defined by

$$< f, g > = \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} f(\theta, \varphi)g(\theta, \varphi)d\Omega = \int_{S^2} f(\theta, \varphi)g(\theta, \varphi)d\Omega = \int_{S^2} fg d\Omega \qquad (2)$$

where $d\Omega = \sin(\theta)d\varphi d\theta$. With respect to the inner product, the spherical harmonics satisfy the orthonormal condition: $\int_{S^2} Y_{lm} Y_{l'm'} d\Omega = \delta_{ll'}\delta_{mm'}$, where δ_{ij} is the Kroneker's delta.

The three functions are independently expressed in terms of the spherical harmonic as: $x(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^x Y_{lm}(\theta, \varphi)$, $y(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^y Y_{lm}(\theta, \varphi)$ and $z(\theta, \varphi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^z Y_{lm}(\theta, \varphi)$. L is the maximal degree of the representation, which determines the degree to which data are smoothed. As we know the values of each function in a sample of points $\{(\theta_i, \varphi_i)\}_{i=1}^n$, the coefficients can be estimated by least squares. In the case of $x(\theta, \varphi)$ (analogously for the other two functions), let $\mathbf{x} = \{x(\theta_i, \varphi_i)\}_{i=1}^n$ be the vector of observations, $\mathbf{c}^x$ the vector containing the coefficients c_{lm}^x and $\mathbf{Y} = \{Y_{lm}(\theta_i, \varphi_i)\}_{i=1}^n$ the matrix of basis function values at the observation points, then $\mathbf{c}^x = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x}$. If the size of the linear equation is extremely large, the coefficients can be also estimated in a least squares fashion by the iterative residual fitting (IRF) algorithm [14]. Finally, a vector-valued function can be built $F(\theta, \varphi) = (x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi))' = \sum_{l=0}^{L} \sum_{m=-l}^{l} \mathbf{c}_{lm} Y_{lm}(\theta, \varphi)$, where $\mathbf{c}_{lm} = (c_{lm}^x, c_{lm}^y, c_{lm}^z)'$.

As in other works that used SPHARM for simple surfaces, a small L results in an acceptable degree of smoothing for this kind of subcortical structures. For more complex structures, such as cortical surfaces, a higher degree (for example, 52) is necessary for a good representation [14]. After inspecting the representations for different values of L, $L = 15$ was visually chosen for our hippocampi. Figure 2 shows the spherical harmonic representations of a hippocampus surface for different L. Note that if L is too small, we miss important aspects of the surface, but with large L we not only fit data but also noise. Hence there was an inevitable trade-off between these two factors in choosing $L = 15$. In order to check if the analysis is robust to the choice of different values of L, the classification results for different values of L are shown in the supplementary material.

Sometimes, a registration or alignment is then carried out. However, in this case it is not necessary as the location was removed previously by translating each hippocampus to the same point in such a way that its centroid coincides with that point ((25,25,25) in our case). Note that the centre of the sphere for the surface parametrization is on each hippocampal

Statist. Med. **0000**, 00 1–12
Prepared using **simauth.cls**

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org

3

centroid, as in Chung et al. [17]. Furthermore, all the hippocampi have the same orientation, so no rotation is needed. As size information (hippocampal volume is a usual discriminatory feature) is important, no scaling correction is carried out. So we analyze the form, which combines the shape and the size information, otherwise scale can be removed by dividing through the centroid's size at the beginning, as in Epifanio and Ventura-Campos [25]. Here, no further alignment is necessary, as in Chung et al. [17], since the coordinates $((\theta, \varphi))$ on two surfaces are corresponding pairs, and therefore the coefficients match each other. For other kind of the surface parametrization, an alignment could be necessary as explained in L. Shen and McPeek [15], where landmarks are used for registration.

In this paper the arguments are angles, but when the argument is time, functions usually exhibit two kind of variation: amplitude and phase variation. The fist one accounts for the size of the shape features in the functions, whereas the second one refers to the location of the features. In case that we had phase variation, the algorithm combining registration with principal components analysis in Kneip and Ramsay [37] could be used, where decomposition of functional variation into amplitude and phase partitions is defined.

2.4. Functional discriminant analysis

Linear discriminant analysis can be used with functions, but regularization is necessary to give meaningful results [22, ch. 8]. Recent advances in functional data classification have been reported by various authors. Many of them involve a type of preprocessing (sometimes implicit) of the functional data (see [38] for a comparison of different methods for univariate functions, and [25] for multivariate functions with one argument). One possible regularization approach is to concentrate on the first few principal components as in [22, ch. 8], or some other finite-dimensional representation of the data, as ICA, which has given better results than PCA and other alternatives in previous literature [25]. So, once the hippocampi are represented in a basis (SPHARM), we can carry out the functional data analysis, beginning with exploring the hippocampal variability by PCA and ICA, and using them for the discriminant analysis.

2.5. Functional PCA (FPCA)

For studying the main sources of variation among the hippocampi, principal component analysis is used. In order to see how PCA works in the functional context, let us recall PCA for Multivariate Data Analysis (MDA). Shortly, summations change into integrations. In MDA, principal components are obtained by solving the eigenequation

$$\mathbf{V}\xi = \rho\xi, \tag{3}$$

where $\mathbf{V}$ is the sample variance-covariance matrix, $\mathbf{V} = (N-1)^{-1}\mathbf{X}'\mathbf{X}$, $\mathbf{X}$ is the centered data matrix, N is the number of subjects observed, and $\mathbf{X}'$ indicates the transpose of $\mathbf{X}$. Moreover, ρ and ξ are an eigenvalue and an eigenvector of $\mathbf{V}$, respectively.

In the functional version of PCA, PCs are replaced by functions. Before analyzing our multivariate functional data with multiple arguments, let us introduce the functional univariate case. Let $\{x_1(t), \ldots, x_N(t)\}$ be the set of univariate observed functions with one scalar argument t. The mean function is defined as the average of the functions point-wise across replications $(\bar{x}(t) = N^{-1}\sum_{i=1}^{N} x_i(t))$. If data have been centered (the mean function has been subtracted), the covariance function $v(s,t)$ is defined analogously by $v(s, t) = (N-1)^{-1}\sum_{i=1}^{N} x_i(s)x_i(t)$. As explained in Ramsay and Silverman [19, Chapter 8], the functional counterpart of equation 3 is the following functional eigenequation

$$\int v(s,t)\xi(t)dt = \rho\xi(s), \tag{4}$$

where ρ is still an eigenvalue, but where $\xi(s)$ is an eigenfunction of the variance-covariance function, rather than an eigenvector. Now, the principal component score corresponding to $\xi(s)$ is computed by using the inner product for functions: $s_i = \int x_i(s)\xi(s)ds$. Note that for multivariate data, the index s is not continuous, but a discrete index j replaces it: $s_i = \sum_j x_{ij}\xi_j$.

For solving the eigenequation 4, the original functions could be discretized. However, we will work with the coefficients of the functions expressed as a linear combination of known basis functions. If the basis is orthonormal, FPCA reduces to the standard multivariate PCA of the coefficient array, as explained in Ramsay and Silverman [19, Sec. 8.4.2], where computational methods for FPCA are reviewed. This reduces the amount of information generated.

With regard to the number of PCs that can be computed, let us note that in the functional context, "variables" now correspond to values of t, and there is no limit to these. Therefore, a maximum of $N - 1$ components can be computed. However, if the number of basis functions M representing the functions is less than N, M would be the maximum.

2.5.1. FPCA with multiple functions and multiple arguments
Let $\{F_i(\theta, \varphi)\}_{i=1}^{N}$ be the set of observed functions. Each F_i consists of three functional data with two arguments representing one hippocampus

$((x_i(\theta,\varphi), y_i(\theta,\varphi), z_i(\theta,\varphi)))$. Three mean functions $(\bar{x}(\theta,\varphi), \bar{y}(\theta,\varphi), \bar{z}(\theta,\varphi))$ and three covariance functions $(v_{XX}((\theta,\varphi),(\vartheta,\phi)), v_{YY}((\theta,\varphi),(\vartheta,\phi)), v_{ZZ}((\theta,\varphi),(\vartheta,\phi)))$ can be computed pointwisely as before for each kind of function, respectively. We can calculate the cross-covariance function of the centered data by (analogously for the combination XZ and YZ) $v_{XY}((\vartheta,\phi),(\theta,\varphi)) = (N-1)^{-1}\sum_{i=1}^{N} x_i(\vartheta,\phi)y_i(\theta,\varphi)$.

An inner product on the space of vector-valued functions is defined by summing the inner products of the components (defined in 2) as

$$< F_1, F_2 >=< x_1, x_2 > + < y_1, y_2 > + < z_1, z_2 > . \tag{5}$$

A typical PC is defined by a three-vector $\xi=(\xi_X, \xi_Y, \xi_Z)$ of weight functions. Now, the PC score for the i-th function is computed by $s_i =< F_i, \xi >= \int_{S^2} x_i\xi_X d\Omega + \int_{S^2} y_i\xi_Y d\Omega + \int_{S^2} z_i\xi_Z d\Omega$. PCs are solutions of the eigenequation system $V\xi = \rho\xi$, which in this case can be written as

$$\begin{array}{l}
\int_{S^2} v_{XX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{XY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega + \int_{S^2} v_{XZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_X(\vartheta,\phi) \\
\int_{S^2} v_{YX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{YY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega + \int_{S^2} v_{YZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_Y(\vartheta,\phi) \\
\int_{S^2} v_{ZX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{ZY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega + \int_{S^2} v_{ZZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_Z(\vartheta,\phi).
\end{array} \tag{6}$$

To solve the eigenequation system, each function in the vector-function F_i is replaced by a vector of basis coefficients, and a single vector is built by joining them together. Then, if $\mathbf{c}^i = (\{c_{ilm}^x\}, \{c_{ilm}^y\}, \{c_{ilm}^z\})$ is that vector of coefficients for F_i, with $l = 0, ..., L$ and $m = -l$ to l, a matrix $\mathbf{C}$ with N rows (one per individual) can be built stacking those vectors. As spherical harmonics are orthonormal, we only need to compute the PCA of $\mathbf{C}$. When PCs have been computed, we separate the parts belonging to each coordinate, as explained in Ramsay and Silverman [19, Sec. 8.5.1] for bivariate FPCA with one argument. Note that the computation is reduced when we work with the coefficients instead of using a lot of variables obtained by discretizing the original functions in a fine grid. With $L = 15$ we have 256 coefficients per coordinate, a total of 756 coefficients (variables). However, for a grid of 2562 vertices on the sphere, the number of variables rises to 7686.

Each eigenvalue ρ divided by the sum of all eigenvalues gives the proportion of variance explained by each eigenfunction, as in the multivariate case. Furthermore, for the j-th principal component $\xi^j = (\xi_X^j, \xi_Y^j, \xi_Z^j)$, the variation accounted for each coordinate can be computed by $< \xi_X^j, \xi_X^j >$, $< \xi_Y^j, \xi_Y^j >$ and $< \xi_Z^j, \xi_Z^j >$ respectively, because their sum is one by definition.

2.6. Functional ICA (FICA)

ICA was successfully used for the classification of univariate functions in Epifanio [38], where was compared with classical and the most recent advances in univariate functional data classification giving results better than or similar to those obtained using the previous techniques in three different problems. Concretely, the proposed descriptors were compared with the methodology introduced in Hastie et al. [39], in Ferraty and Vieu [40] including the multivariate partial least-squares regression (MPLSR) method in its semi-metric and PCA, in Rossi and Conan-Guez [41], in Ferré and Villa [42], in Rossi and Villa [43], and as Li and Yu [44] use the same example, we can also compare the results in Epifanio [38] with those of Li and Yu [44] (see Epifanio [38] for details). That methodology was extended to the multivariate case with one argument in Epifanio and Ventura-Campos [25], where the best discriminant results were obtained with the ICA coefficients compared with FPCA and other alternatives (the penalized discriminant analysis proposed by Hastie et al. [39] and the nonparametric curve discrimination method with the semi-metric based on FPCA and MPLSR introduced by Ferraty and Vieu [40]) in the functional approach, and the set and landmark approach. Here we extend the methodology for multivariate functions with two (or more) arguments, as if ICA has been useful with the classification of functions with one argument, the same can happen with more general functions.

Let us recall ICA for MDA. Assume that the data matrix $\mathbf{X}$ is a linear combination of non-Gaussian (independent) components i.e. $\mathbf{X} = \mathbf{SA}$ where columns of $\mathbf{S}$ contain the independent components and $\mathbf{A}$ is a linear mixing matrix. ICA attempts to "un-mix" the data by estimating an un-mixing matrix $\mathbf{W}$ with $\mathbf{XW} = \mathbf{S}$. Under this generative model, the measured "signals" in $\mathbf{X}$ will tend to be "more Gaussian" than the source components (in $\mathbf{S}$) due to the Central Limit Theorem. Thus, in order to extract the independent components or sources, we search for an un-mixing matrix $\mathbf{W}$ that maximizes the non-gaussianity of the sources.

For univariate functions, assume that we observe N linear mixtures $x_1(t), ..., x_N(t)$ of K independent components $s_j(t)$: $x_i(t) = \sum_{j=1}^{K} a_{ij}s_j(t)$, for all i. Each pair $s_j(t)$ and $s_k(t)$, at each time instant t, are statistically independent. In practice, we have discretized functions ($\mathbf{x}_i = \{x_i(t_k); k = 1, ..., p\}$), therefore we can consider the $p \times N$ data matrix $\mathbf{X} = \{x_i(t_k)\}$.

However, unlike our previous works with one argument, instead of discretizing the functions we will work with the coefficients in a functional basis for reducing the computational burden. Suppose that each function has basis expansion: $x_i(t) = \sum_{m=1}^{M} b_{im}G_m(t)$. If we define x a vector-valued function with components $x_1, ..., x_N$, and G the vector-valued function with components $G_1, ..., G_M$, we can express the simultaneous expansion of all N functions as: $x = \mathbf{B}G$, where $\mathbf{B}$ is the coefficient matrix, with size $N \times M$. If we perform ICA on $\mathbf{B}'$, we obtain $\mathbf{B}' = \mathbf{S_b}\mathbf{A_b}$, so we can consider x

Statist. Med. **0000**, 00 1–12
Prepared using *simauth.cls*

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org 5

$= \mathbf{B}G = \mathbf{A'_b}\mathbf{S'_b}G$, i.e. the observed data x are generated by a process of mixing the K components $I = \mathbf{S'_b}G$ (rows of $\mathbf{S'_b}$ contain the independent components). The expansion of any function $\tilde{x}(t)$ not included in the original x in terms of these ICA components will be of the form: $\tilde{x}(t) = \sum_{j=1}^{K} \tilde{a}_j I_j(t)$, with $I_j(t)$ the j-th component of I. If we estimate I and G in p points ($\{t_k; k = 1, ..., p\}$), we can build the $p \times K$ matrix $\mathbf{I}$ and the $p \times M$ matrix $\mathbf{G}$, and hence $\mathbf{I} = \mathbf{GS_b}$. The K-vector $\tilde{\mathbf{a}}$ containing the coefficients $\tilde{a}_j$ can be easily obtained by least squares fitting [19]: $\tilde{\mathbf{a}} = (\mathbf{I'I})^{-1}\mathbf{I'}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = \{\tilde{x}(t_k)\}_{k=1}^{p}$. This yields $\tilde{\mathbf{a}} = (\mathbf{S'_b}\mathbf{G'}\mathbf{GS_b})^{-1}\mathbf{S'_b}\mathbf{G'}\tilde{\mathbf{x}}$. Analogously, for the G basis, $\tilde{x}(t) = \sum_{m=1}^{M} \tilde{b}_m G_m(t)$, and the M-vector $\tilde{\mathbf{b}}$ containing the coefficients $\tilde{b}_m$ can be computed as: $\tilde{\mathbf{b}} = (\mathbf{G'G})^{-1}\mathbf{G'}\tilde{\mathbf{x}}$. When the basis G is orthonormal, meaning that $\mathbf{G'G}$ is the identity matrix :

$$\tilde{\mathbf{a}} = (\mathbf{S'_b}\mathbf{G'}\mathbf{GS_b})^{-1}\mathbf{S'_b}\mathbf{G'}\tilde{\mathbf{x}} = (\mathbf{S'_b}\mathbf{S_b})^{-1}\mathbf{S'_b}\tilde{\mathbf{b}}. \tag{7}$$

If the functions have more than one argument, the discussion is identical. When having multivariate functional data, we can concatenate the coefficients for each function into a single long vector, as done in Sec. 2.5.1 for computing multivariate FPCA. In our case, $\tilde{\mathbf{b}}$ would be $(\mathbf{c}^i)'$.

Before the application of the ICA algorithm, it is useful to reduce the dimension of the data previously by PCA (for details, see Hyvärinen et al. [45, Section 5]), thus reducing noise and preventing overlearning [46, Section 13.2]. Therefore we compute the PCA first, retaining a certain number of components, and then estimate the same number of independent components as the PCA reduced dimension.

2.6.1. *Functional linear discriminant*
Functional linear discriminant can be used if the objective is also to discriminate between different groups and to understand the way in which these groups differ. The coefficients ($\tilde{\mathbf{a}}$) for ICA components will constitute the feature vector used for the classification step, as made in Epifanio [38] for univariate functions and Epifanio and Ventura-Campos [25] for multivariate functions with one argument. The scores for functional PCs can also be used, although in Epifanio and Ventura-Campos [25] the results were not so good as for ICA. The use of PCA is quite common before the classifier is applied, such as in Beg et al. [18] or Shen et al. [13] (although they did not applied PCA on the coefficients but on the landmarks: points estimated on the surfaces).

We propose to compute a linear discriminant vector function $\lambda^j(\theta, \varphi) = (\lambda_X^j(\theta, \varphi), \lambda_Y^j(\theta, \varphi), \lambda_Z^j(\theta, \varphi))$ based on FICA as done in Ramsay and Silverman [22, Chapter 8] with FPCA. This function $\lambda^j(\theta, \varphi)$ would be the functional counterpart of the linear discriminant or canonical variate [47, Chapter 3], therefore, $d_i^j = < F_i, \lambda^j > = \int_{S^2} x_i \lambda_X^j d\Omega + \int_{S^2} y_i \lambda_Y^j d\Omega + \int_{S^2} z_i \lambda_Z^j d\Omega$ would return the score or discriminant value of F_i. If we express both functions in the spherical harmonics base, due to the orthonormality, d_i^j is just the inner product of two vectors, $d_i^j = (\lambda^j)(\mathbf{c}^i)'$, where λ^j is the vector with the coefficients of $\lambda^j(\theta, \varphi)$ in that base. In this way, the problem is reduced to find these coefficients in the spherical harmonic expansion.

Assume that there are Q groups, each of them with size N_i ($\sum_{i=1}^{Q} N_i = N$) and we apply the standard linear discriminant analysis (LDA) to the $K \times N$ matrix $\mathbf{A}$ with the coefficients of the K ICA components. This yields a $K \times r$ matrix $\mathbf{L}$ ($r = \min\{K, Q - 1\}$ is the number of discriminant functions) giving the $r \times N$ matrix $\mathbf{D}$ of discriminant values ($\mathbf{D} = \mathbf{L'A}$). By equation 7, $\mathbf{A} = (\mathbf{S'_b}\mathbf{S_b})^{-1}\mathbf{S'_b}\mathbf{C'}$, where $\mathbf{S_b}$ is the $3M \times K$ matrix containing the independent components of $\mathbf{C'}$, the $N \times 3M$ matrix with the coefficients in the spherical harmonics base with $L = 15$ (hence $M = 256$). As we had $\mathbf{D} = \boldsymbol{\Lambda}\mathbf{C'}$ where $\boldsymbol{\Lambda}$ is the $r \times 3M$ matrix with the coefficients of the r functions $\lambda^j(\theta, \varphi)$ ($j = 1, ..., r$) in the spherical harmonic base (λ^j is the j-th row), then $\boldsymbol{\Lambda} = \mathbf{L'}(\mathbf{S'_b}\mathbf{S_b})^{-1}\mathbf{S'_b}$.

As the problem has been reduced to a MDA problem (although the basis choice plays a key role), we can consider significance tests under the assumption of multivariate normality of the coefficients in $\mathbf{A}$ [48, Sec. 8.6] (ICA looks for non-gaussianity in $\mathbf{S}$ not in $\mathbf{A}$).

2.7. *Visualization of the results*

In order to display the effect of each functional PC, FICA component or discriminant function, a small set of suitable multiples (positive or negative) of the function in question is added to the the mean function (mean hippocampus), which can be displayed for each multiple separately. This procedure is usual in shape and FDA literature [22]. Furthermore, a vector map can be plotted: vectors can be drawn from the mean shape to the surface formed by the mean plus the multiple of the function in question. We can also color the mean hippocampus using the magnitude (norm) of those vectors.

Coefficients in FICA or in FPCA base (scores) and the discriminant values could be also plotted.

3. Results

The main code (mostly in Matlab) and data are available at http://www3.uji.es/~epifanio/RESEARCH/alzfda.rar. Two valuable packages are: the SurfStat package (http://www.math.mcgill.ca/keith/surfstat) and its extension (http://www.stat.wisc.edu/~mchung/research/amygdala/) [17] and the FastICA package [49]. The FastICA algorithm (which includes the PCA computation) is used for obataining ICA. Although we have not considered other algorithms for obtaining ICA, there is no restriction for using any other algorithm, but FastICA is an efficient and popular algorithm. It is based on a fixed-point iteration scheme maximizing non-Gaussianity as a measure of statistical independence. We have used the default parameters of FastICA for obtaining ICA: the independent components are estimated one-by-one (not in parallel), the nonlinearity used in the fixed-point algorithm is a Gaussian function, and the default parameters for controlling the convergence (see the software [49] for details).

We have only considered the left hippocampi as in Beg et al. [18] for illustrating and assessing the proposed methodology. The database is quite small for obtaining valid medical conclusions, although the methodology could be used without modification with a larger database for the left and right hippocampi. The left hippocampal volume has been shown to be better at discriminating MCI status [50]. The volume can be estimated as the sum of the slice areas, i.e. the number of pixels belonging to each segmented hippocampal slice. Nevertheless, the numerical results for the right hippocampi are shown in the supplementary material.

3.1. Functional approach: FPCA and FICA

FPCA is carried out to describe the variability. The first three principal components explain 49.48% of the whole variance, made up of 24.73%, 14.15% and 10.60% respectively. 91.27% of the variability is explained by the first sixteen components. Figure 3 presents visual representations of the shape variation along the first three principal components. Figures in the supplementary material can help to interpret them. The first component correspond to a size component, mostly concentrated on the head and the tail, but also in the body. 45.25% of the variation in this component is due to the x-coordinates (27.17% and 27.58% due to the y and z-coordinates). Component 2 is focused on a part of the tail. In this component, 20.37% of variability comes from the y-coordinates, and the rest is divided between the x and z-coordinates. Finally, the third component is concentrated almost entirely on the whole tail. The proportion of the variability in this component is 41.81% for the y-coordinates (23.51% and 34.67% due to the x and z-coordinates).

It is interesting to distinguish between patients with AD, MCI and elderly CN subjects, however in the literature it is quite common to consider the pair-wise comparisons among the three groups. The important CN-MCI subproblem is analyzed here, together with the numerical results for the three group problem. For a detailed analysis of the other problems, see the supplementary material.

In order to test the ability to classify the subjects into their correct group, cross-validation is performed using leave-one-out trials. In each trial, one subject is set aside and FPCA is performed on the remaining subjects (the training set). On the one hand, the LDA classifier is trained with the first J PC scores of the training set. On the other hand, we compute (and record) the leave-one-out (LOU) prediction error for that training set. The test subject is projected onto the same principal components obtained from training set alone, and classified with the trained LDA. This predicted class is preserved to produce the LOU estimate of the correct classification percentage for J components, since this process is repeated in turn for each of the subjects. Note that as FPCA is computed with different data, PCs from different iterations are not comparable, especially those with low variance. We have the estimated accuracy for various values of the number J of principal components. In order to select the appropriate number of components, we do not choose the best result, as the accuracy would be too optimistic, but a double or nested cross-validation is done. We consider the recorded LOU prediction errors for each training set, and build a matrix with them. The number of rows is the number of subjects whereas the number of columns is the total number of values J considered. We compute the mean for each column, and the model (value J) selected is that one which gives smallest mean (in case of tie, we select that with less components, the most parsimonious). Hence, the estimated classification accuracy is the estimated accuracy by LOU for the model selected by the nested LOU. Table 2 shows the results for FPCA together with the rest of methods for the CN vs MCI problem. In the supplementary material a table with the mean and standard deviation for each column of the matrix can be seen, and also a table with the external LOU accuracies for different J values.

The same cross-validation strategy considered for FPCA is followed with FICA, whose results are displayed in Table 2. Furthermore, we have computed the functional linear discriminant using all the subjects in the set for the number of components selected ($J = 6$ in this subproblem). In Figure 4 it is visualized as explained in Sec. 2.7. It suggests a small loss in the CA1 and a part of the subiculum in the body of the hippocampus. The discriminant function significantly separates the groups (Wilks' $\Lambda = 0.197$, p-value = 0.002).

Statist. Med. **0000**, 00 1–12
Prepared using **simauth.cls**

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org 7

3.2. Comparative performance

We reproduce the methodology in Chung et al. [17] with our data. We perform multivariate linear modeling [51] on our spherical harmonic representation, testing the effect of group variable in the model. There is no statistically significant shape difference at α=0.05 when we test for group differences at each vertex of the hippocampal surface (see Figure 5 displaying the F-statistic value on the mean hippocampus). Although testing of group mean difference and discriminant analysis are different problems (note that the coefficients for the discriminant function in MDA are derived so as to maximize the differences between the group means), we have included this result for highlighting the utility of the proposed discriminant function since significance maps of group differences usually appear in neuroimaging literature together with classification results [16].

The methodology in Gerardin et al. [16] is applied to our data: the SPHARM coefficients are classified with a support vector machine (SVM). Student's t-tests were used for determining which coefficients best separate the groups, with a bagging strategy (we use the same strategy but with the absolute value of the t statistic for really keeping only those coefficients which are always significantly different since it is a two tailed test). The result obtained with LOU is in Table 2. The number of coefficients is selected by nested LOU. A linear kernel is considered, since better results are obtained than with radial basis functions with different scaling factors.

Recently, Clemmensen et al. [52] have proposed a sparse discriminant analysis (SDA) for the high-dimensional setting (the number of variables is large relative to the number of subjects). This method performs linear discriminant analysis with a sparseness criterion imposed such that classification and feature selection are performed simultaneously. SDA is applied to our SPHARM coefficients. We use the sparseLDA package with default parameters (except for the desired number of coefficients to be selected), which is available from http://www2.imm.dtu.dk/~lhc/. The result obtained with LOU is in Table 2. The number of coefficients is chosen by nested LOU.

Beg et al. [18] proposed four shape features for discriminating CN vs MCI with left hippocampi. They reported the following accuracies for each set of features: 68.1% for 3D moment invariants features, 75% for 3D tensor invariant features, 77.3% for 3D Laplacian invariant features, and 86.3% for 3D geodesic shape invariants features. Note that in their database there are 26 CN and 18 MCI (a proportion similar to our database). Although results should be compared with caution since the databases are different, FPCA and FICA obtain better accuracies.

Other approaches, such as that in Shen et al. [13], evaluate the spherical representation in a serie of parameter locations, obtaining the coordinates on the surface, and work with these landmarks (SPHARM-PDM from Point Distribution Model) for classification. We do not consider this approach since it is well-known that some regularization is necessary in order to obtain meaningful results [19, Ch. 11], [22, Ch. 8]. In fact, in Figure 5, where each spatial location is considered separately, there are a lot of small spots. However, in Figure 4 the discriminative zones are larger and more homogeneous.

Table 3 shows the confusion matrices for each method when the three groups are considered jointly, except for SVM in Gerardin et al. [16], whose methodology is only for problems with two groups. Best or equal accuracies are achieved with the functional approach. The CN group is correctly identified by FPCA, FICA and SDA, and SDA classifies correctly the AD group (there is only one misclassification is this group for the rest of methods, which classified an AD individual as MCI). The main errors are in the MCI group: with volumen only one individual is correctly identified in this group (two individuals are misclassified as CN and three as AD); with SDA two individuals are misclassified as CN and other two as AD, while the less number of errors are achieved with FPCA and FICA, where two individuals are misclassified as CN and one as AD. The two MCI individuals misclassified as CN are the same for all the methods, as well as the misclassified MCI individual as AD for FPCA and FICA.

4. Conclusions

Our novel contribution is to model the 3D hippocampal surfaces with a FDA approach, as hippocampi are in fact functions defined over two spherical angles. This approach allows to carry out the same analysis as any other branch of statistics [19]. As there are few studies with functional data with more than one argument, we have extended the functional data methodology for functions with two arguments (angles in this case), introducing FICA for the first time. FPCA and FICA with two arguments have been used for shape description and classification. A functional linear discriminant with FICA has been also defined. To the best of our knowledge it has been used for the first time in localizing the differences among groups instead of the significance maps, and meaningful results are obtained with it. The functional linear discriminant with FICA has given more details and coherence than that with FPCA (see the supplementary material). Classification results are better with FPCA and FICA (both with identical results), than the other alternatives considered, showing that feature extraction is a powerful method [53, Sec. 5.3.]. The same conclusions are reached with the other subproblem considered in supplementary material. The good results in this important problem are other of our main contributions: in Table 2 higher accuracy, sensitivity and specificity is achieved with the FDA approach (FPCA and FICA) than with

other alternatives, even with a fewer number of features (except for volume which is univariate). We have focused on the analysis of left hippocampi, as they seem to discriminate AD condition better than the right hippocampi (see the supplementary material). As future work, it would be interesting to analyze the hippocampal asymmetry between left and right hippocampus in AD (as made for example with schizophrenia in [54]), using FDA and exploiting the potential of this perspective.

Our sample size is small and we use LOU (with nested LOU for selecting the number of features) for assessing the method. With a larger database could be possible to split the data into a training, validation and test set, and instead of selecting the number of features, the features could be selected [48, Sec. 8.9.], and classification results could be improved.

The FDA approach is not restricted to hippocampi, it could be used with other structures. Furthermore, our methodology could be extended to deal with functional data combined with multivariate random variables. Ramsay and Silverman [19, Ch. 10] defined FPCA of hybrid data (univariate functions with one argument together with a vector). Maybe it is interesting to consider the age, the education years or the intracranial volume, etc.

Finally, other interesting future topic includes using the FDA approach in other problems (such as Functional ANOVA or detection of outliers) with 3D shapes, for example for assessing the evolution along time in the same subjects in a longitudinal study.

Acknowledgments

References

[1] Grundman, M., Petersen, R., Ferris, S., Thomas, R., Aisen, P., Bennet, D., Foster, N., Jack, C., Galasho, D., Dondy, R., Kaye, J., and Sano, M. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology*, 61:59–66, 2004.

[2] Petersen, R. C. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256:183–194, 2004.

[3] Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Boeve, B. F., Waring, S. C., Tangalos, E. G., and Kokmen, E. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, 52:1397–1403, 1999.

[4] Mungas, D., Jagust, W. J., Reed, B. R., Kramer, J. H., Weiner, M. W., Schuff, N., Norman, D., Mack, W. J., Willis, L., and Chui, H. C. MRI predictors of cognition in subcortical ischemic vascular disease and Alzheimer's disease. *Neurology*, 57:2229–2235, 2001.

[5] Styner, M., Lieberman, J. A., Pantazis, D., and Gerig, G. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical Image Analysis Journal*, 8(3):197–203, 2004.

[6] Styner, M., Gerig, G., Joshi, S. C., and Pizer, S. M. Automatic and robust computation of 3D medial models incorporating object variability. *International Journal of Computer Vision*, 55(2-3):107–122, 2003.

[7] Golland, P., Grimson, W. E. L., Shenton, M. E., and Kikinis, R. Deformation analysis for shape based classification. In *Proceedings of the 17th International Conference on Information Processing in Medical Imaging*, IPMI '01, pages 517–530, 2001. ISBN 3-540-42245-5.

[8] Joshi, S. C., Miller, M. I., and Grenander, U. On the geometry and shape of brain sub-manifolds. *Internation Journal of Pattern Recognition and Artificial Intelligence*, 11(8):1317–1343, 1997.

[9] Park, Y., Priebe, C. E., Miller, M. I., Mohan, N. R., and Botteron, K. N. Statistical analysis of twin populations using dissimilarity measurements in hippocampus shape space. *Journal of Biomedicine and Biotechnology*, 2008(Article ID 694297, doi:10.1155/2008/694297):5 pages, 2008.

Statist. Med. **0000**, 00 1–12
Prepared using simauth.cls

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org　**9**

[10] Shen, K., Fripp, J., Meriaudeau, F., Chetelat, G., Salvado, O., and Bourgeat, P. Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models. *NeuroImage*, 59(3):2155–2166, 2012.

[11] Gerig, G., Styner, M., Jones, D., Weinberger, D., and Lieberman, J. Shape analysis of brain ventricles using SPHARM. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01)*, pages 171–178, 2001.

[12] Gu, X., Wang, Y., Chan, T. F., Thompson, P. M., and Yau, S. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23(8):949–958, 2004.

[13] Shen, L., Ford, J., Makedon, F., and Saykin, A. A surface-based approach for classification of 3D neuroanatomic structures. *Intelligent Data Analysis*, 8(6):519–542, 2004.

[14] Chung, M. K., Dalton, K. M., Shen, L., Evans, A. C., and Davidson, R. J. Weighted fourier series representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26(4):566–581, 2007.

[15] L. Shen, H. F. and McPeek, M. Modeling 3-dimensional morphological structures using spherical harmonics. *Evolution*, 4(63):1003–1016, 2009.

[16] Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., and Colliot, O. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47 (4):1476–1486, 2009.

[17] Chung, M. K., Worsley, K. J., Nacewicz, B. M., Dalton, K. M., and Davidson, R. J. General multivariate linear modeling of surface shapes using surfstat. *NeuroImage*, 53(2):491–505, 2010.

[18] Beg, M. F., Raamana, P. R., Barbieri, S., and Wang, L. Comparison of four shape features for detecting hippocampal shape changes in early Alzheimers. *Statistical Methods in Medical Research*, published online 30 May 2012(DOI: 10.1177/0962280212448975):1–24, 2012.

[19] Ramsay, J. O. and Silverman, B. W. *Functional Data Analysis*. Springer, 2nd edition, 2005.

[20] Ferraty, F. and Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, 2006.

[21] Ferraty, F. and Romain, Y. *The Oxford Handbook of functional data analysis*. Oxford University Press, 2011.

[22] Ramsay, J. O. and Silverman, B. W. *Applied Functional Data Analysis*. Springer, 2002.

[23] Luo, W., Cao, J., Gallagher, M., and Wiles, J. Estimating the intensity of ward admission and its effect on emergency department access block. *Statistics in Medicine*, To appear(doi: 10.1002/sim.5684), 2012.

[24] Zhou, Y. and Sedransk, N. A new functional data-based biomarker for monitoring cardiovascular behavior. *Statistics in Medicine*, 32(1):153164, 2013.

[25] Epifanio, I. and Ventura-Campos, N. Functional data analysis in shape analysis. *Computational Statistics & Data Analysis*, 55(9):2758–2773, 2011.

[26] Stoyan, D. and Stoyan, H. *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics*. Wiley, 1994.

[27] Watson, C., Andermann, F., Gloor, P., Jones-Gotman, M., Peter, T., A., E., Olivier, A., Melanson, D., and G., L. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. *Neurology*, 42(9):1743–1750, 1992.

[28] Hasboun, D., Chantôme, M., Zouaoui, A., Sahel, M., Deladoeuille, M., Sourour, N., Duyme, M., Baulac, M., Marsault, C., and Dormont, D. MR determination of hippocampal volume: Comparison of three methods. *American Journal of Neuroradioly*, 17:1091–1098, 1996.

[29] Kindratenko, V. V. On using functions to describe the shape. *Journal of Mathematical Imaging and Vision*, 18: 225–245, 2003.

[30] Timsari, B. and Leahy, R. M. Optimization method for creating semi-isometric flat maps of the cerebral cortex. In *Proc. SPIE, Medical Imaging*, volume 3979, pages 698–708, 2000.

[31] Brechbühler, C., Gerig, G., and Kübler, O. Parametrization of closed surfaces for 3-D shape description. *Computer Vision and Image Understanding*, 61(2):154–170, 1995.

[32] Macdonald, D., Kabani, N., Avis, D., and Evans, A. C. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12(3):340–356, 2000.

[33] Alfeld, P., Neamtu, M., and Schumaker, L. L. Fitting scattered data on sphere-like surfaces using spherical splines. *Journal of Computational and Applied Mathematics*, 73(1-2):5–43, 1996.

[34] He, Y., Li, X., Gu, X., and Qin, H. Brain image analysis using spherical splines. In *Proc. of Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 633–644, 2005.

[35] Nain, D., Styner, M., Niethammer, M., Levitt, J. J., Shenton, M., Gerig, G., and Tannenbaum, A. Statistical shape analysis of brain structures using spherical wavelets. In *Proceedings of the Fourth IEEE International Symposium on Biomedical Imaging*, page 209–212, 2007.

[36] Yu, P., Grant, P. E., Qi, Y., Han, X., Ségonne, F., Pienaar, R., Busa, E., Pacheco, J., Makris, N., Buckner, R. L., Golland, P., and Fischl, B. Cortical surface shape analysis based on spherical wavelets. *IEEE Transactions on Medical Imaging*, 26(4):582 – 597, 2007.

[37] Kneip, A. and Ramsay, J. O. Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165, 2008.

[38] Epifanio, I. Shape descriptors for classification of functional data. *Technometrics*, 50(3):284–294, 2008.

[39] Hastie, T., Buja, A., and Tibshirani, R. Penalized discriminant analysis. *Annals of Statistics*, 23:73–102, 1995.

[40] Ferraty, F. and Vieu, P. Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis*, 44:161–173, 2003.

[41] Rossi, F. and Conan-Guez, B. Functional multi-layer perceptron: a nonlinear tool for functional data analysis. *Neural Networks*, 18(1):45–60, 2005.

[42] Ferré, L. and Villa, N. Multilayer perceptron with functional inputs: an inverse regression approach. *Scandinavian Journal of Statistics*, 33(4):807–823, 2006.

[43] Rossi, F. and Villa, N. Support vector machine for functional data classification. *Neurocomputing*, 69(7–9):730–742, 2006.

[44] Li, B. and Yu, Q. Classification of functional data: A segmentation approach. *Computational Statistics and Data Analysis*, 52:4790–4800, 2008.

[45] Hyvärinen, A., Karhunen, J., and Oja, E. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[46] Hyvärinen, A., Karhunen, J., and Oja, E. *Independent component analysis*. Wiley, New York, 2001.

[47] Ripley, B. D. *Pattern recognition and neural networks*. Cambridge University Press, 1996.

[48] Rencher, A. C. *Methods of Multivariate Analysis*. Wiley, 2nd edition, 2002.

[49] Hyvärinen, A. Fast and robust fixed-point algorithms for indepent component analysis, http://www.cis.hut.fi/projects/ica/fastica/. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[50] Müller, M. J., Greverus, D., Weibrich, C., Dellani, P. R., Scheurich, A., Stoeter, P., and Fellgiebel, A. Diagnostic utility of hippocampal size and mean diffusivity in amnestic MCI. *Neurobiology of Aging*, 28(3):398–403, 2007.

[51] Taylor, J. and Worsley, K. Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, 36(1):1–27, 2008.

[52] Clemmensen, L., Hastie, T., Witten, D., and B., E. Sparse discriminant analysis. *Technometrics*, 53(4):406–413, 2011.

Statist. Med. **0000**, 00 1–12
Prepared using *simauth.cls*

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org   **11**

[53] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning. Data mining, inference and prediction.* Springer-Verlag, second edition, 2009.

[54] Wang, L., Joshi, S., Miller, M., and Csernansky, J. Statistical analysis of hippocampal asymmetry in schizophrenia. *Neuroimage*, 14(3):531–45, 2001.

5. Figures and Tables

Table 1. Description of the database. The number inside the parentheses is the standard deviation.

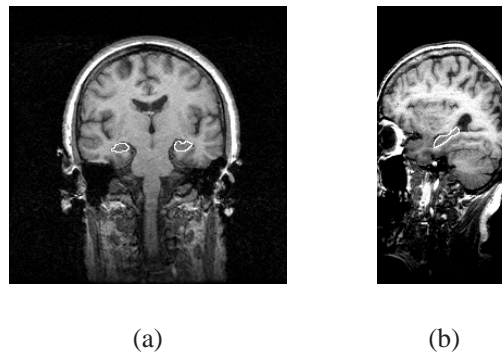|	CN, $N_1 = 12$	MCI, $N_2 = 6$	AD, $N_3 = 10$
Sex (Male/Female)	5/7	2/4	1/9
Mean age	70.17 (3.43)	75.50 (3.33)	71.5 (4.35)
Mean CDR (Clinical Dementia Rating)	0 (0)	0.5 (0)	0.95 (0.15)

(a)   (b)

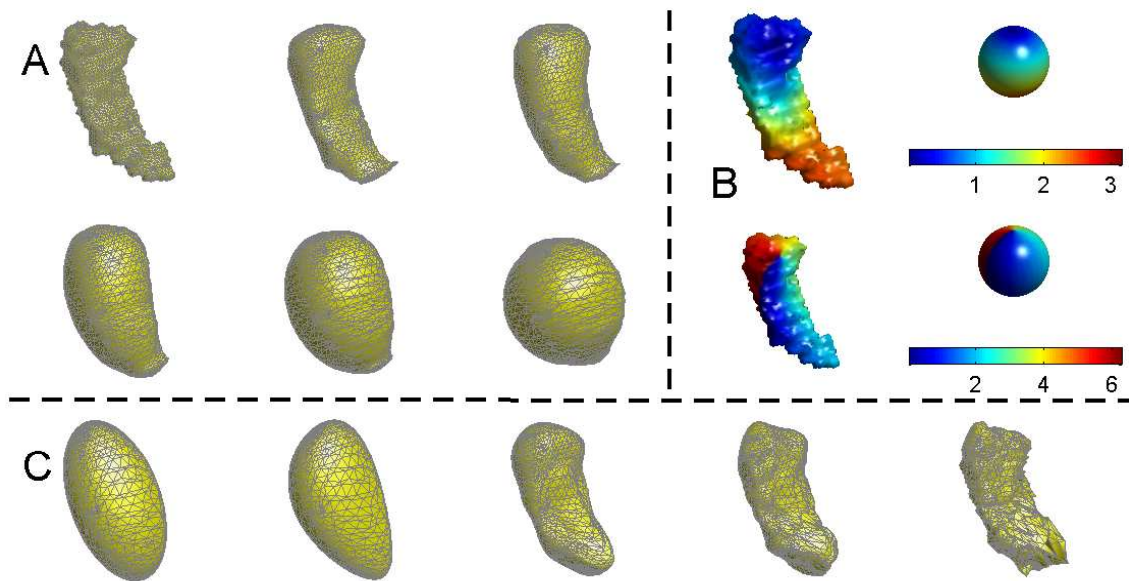**Figure 1.** Hippocampal outlines in a coronal (a) and sagittal (b) slice.

Figure 2. For one hippocampus: (A) the surface flattening process from the original one in the top left of the figure to the bottom right. The same level sets as in [17] are used (1.0, 0.6, 0.2, -0.2, -0.6, -1.0). (B) The spherical angles projected onto the hippocampus surface (for θ and φ respectively), together with a unit sphere showing the angles. (C) The spherical harmonic representations of the hippocampus surface for $L = 1, 2, 8, 15, 30$ (from left to right).

Figure 3. The effect of the first PC (first row), the second PC (second row) and the third PC (third row) on the mean shape: from -2 (left) to 2 (right) in intervals of 1 standard deviations for each component.

Table 2. Performance for the CN vs MCI for the different methods, selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the accuracy.

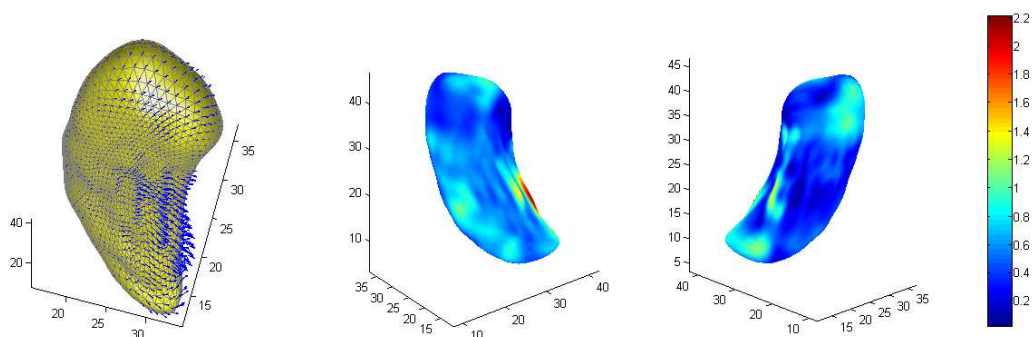| Method	Volume	FPCA	FICA	SVM [16]	SDA
Accuracy (%)	83.33	88.89	88.89	77.78	72.22
No. features	1	5	6	12	11
Sensitivity (%)	66.67	66.67	66.67	50	33.33
Specificity (%)	91.67	100	100	91.67	91.67

Figure 4. Functional linear discriminant with FICA for CN vs MCI (from left to right): vector map with the magnitude and directionality, and magnitude for two views. The arrows show the direction in which the discriminant score increases most rapidly. The color shows the norm of those arrows.
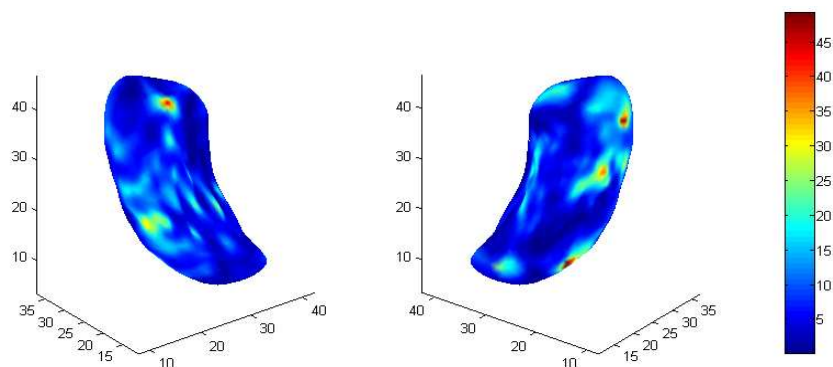
Statist. Med. **0000**, 00 1–12
Prepared using *simauth.cls*

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org 13

Figure 5. Two views of the F-statistic map of shape difference between CN and MCI (the random field based thresholding corresponding to α=0.05 level is 69.11, for α=0.1 level is 59.17, while the maximum F-statistic value is 49.02). F-statistic does not achieve to make clear the differences as the Figure 4.

Table 3. Confusion matrices (in absolute counts) for the CN, MCI and AD for the different methods (Volume/ FPCA / FICA / SDA), selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the entries. The accuracies for each method are (in parentheses the No. features): Volume 75.00% (1), FPCA 85.71% (8), FICA 85.71% (9), SDA 85.71% (13).

	CN	MCI	AD
CN	11 / 12 / 12 / 12	1 / 0 / 0 / 0	0 / 0 / 0 / 0
MCI	2 / 2 / 2 / 2	1 / 3 / 3 / 2	3 / 1 / 1 / 2
AD	0 / 0 / 0 / 0	1 / 1 / 1 / 0	9 / 9 / 9 / 10

Supplementary material to "Hippocampal shape analysis in Alzheimer's disease using Functional Data Analysis"

Irene Epifanio[a]*, Noelia Ventura-Campos[b]

1. Introduction

This Supplementary Material contains the analysis of the AD patients versus CN and versus MCI and the analysis for the three groups: CN, MCI and patients with early AD.

2. Results

Figure 1 shows the magnitude of the first three principal component over the mean hippocampus for the whole database. The viewpoints have been selected in order to visually appreciate better the effect. As code and data are available at http://www3.uji.es/~epifanio/RESEARCH/alzfda.rar, figures can be reproduced and the view can be interactively rotated. Head, body and tail are the three parts that make up a hippocampus [1]. A schematic representation of the hippocampal subfields is shown in Figure 2, which can help in the interpretation. Analogously, Figure 3 shows the magnitude of the first nine independent components over the mean hippocampus for the whole database. The effect for the first and third components is distributed along the whole hippocampus except the more extreme zones of the tail and head; while the rest of components are concentrated in different parts of the hippocampus: tail (second, seventh, eighth and ninth), subiculum (fifth and sixth), and head (fourth).

Table 1 gives the accuracies for the CN vs MCI problem using different values of L for representing the hippocampi. Note that the results are similar to those obtained with L =15, but when L increases (for L bigger than the chosen L = 15) the performances are a bit worse, maybe because we could be fitting noise. For small L the performances are similar. Curiously with L = 3 the numerical results are a bit better, but it would not be possible to appreciate where the differences are with the functional linear discriminant, since with L = 3 the degree of smoothing is very high (see Figure 2 in the paper).

Table 2 gives the summary analysis of the accuracies for the nested LOU (the recorded LOU prediction errors for each training set) for different values of J in the CN vs MCI subproblem. The biggest mean accuracy for each method corresponds with one of the smallest standard deviation. Note that these accuracies are bigger than those reported in Table 2 of the paper, since these accuracies are computed with each training set, and therefore they are overestimated sometimes substantially [2, ch. 7], while in Table 2 of the paper the external LOU results are reported (each test subject is a completely new data not used in the analysis or the selection of the model, for ascertaining the generalization capability of the classifier). In Table 3 the accuracies for the external LOU for different values of J are shown (the selected ones by the nested LOU, whose performances appear in Table 2 of the paper, are in a frame box).

[a] *Dept. Matemàtiques, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló, Spain*
[b] *Dept. Psicologia Bàsica, Clínica i Psicobiologia, Universitat Jaume I, Spain*
Correspondence to: Tel.: +34-964728390, fax: +34-964728429. E-mail: epifanio@uji.es

Figures 4, 5 and 6 display two views of the functional linear discriminant with FICA using all the subjects in each subset for the number of components selected by nested LOU. Considering all these figures jointly, together with Figure 3 in the paper, the evolution of the disease (where it begins and where it ends) can be seen perfectly. As said in the paper, it suggests a small loss in the CA1 and a part of the subiculum in the body of the hippocampus for CN vs MCI. In MCI vs AD, there is a larger difference in CA1, with a bigger value in the tail. Hence, the loss would begin from the head to the tail. Furthermore, there are many differences along the whole subiculum, no so localized as in the CN vs MCI case. In summary, it is clear that the disease would begin in the CA1 region from the head to the tail, with a progressive loss in the subiculum, although this loss is not so pronounced as in CA1.

Figure 7 shows the functional linear discriminant with FPCA defined by [3, ch. 8] for the different subproblems. The viewpoints have been selected in order to visually appreciate better the effect. Note that the discriminating ability for the discriminant functions in the MCI vs AD and CN vs AD subproblems is in doubt since their respective p-values are higher than $\alpha = 0.05$. When the three groups are considered jointly, we obtain the results expected in the hippocampal head (CA1 subregion), and the tail of the hippocampus with a similar level of discrimination, and a loss in the subiculum in the hippocampal body, although not so pronounced. As regards the subproblems CN vs MCI and CN vs AD, we obtain the areas that were expected: the head the hippocampus (CA1 subregion) and the tail of the hippocampus, with greater discrimination in the CN vs MCI subproblem. This is not coherent since the AD patients have a greater atrophy or hippocampal volume loss, so it would be expected a higher value of discrimination between CN vs AD. For the MCI vs AD subproblem, the same configuration (discriminant regions) that for FICA is obtained. However, the discriminant level is equal for the more discriminant zones (CA1 subregion in the front of the head and the tail of the hippocampus), whereas with FICA it can be observed not only a high discrimination in the CA1 subregion located in the front of the head, but also along the body of the hippocampus and a bigger difference in the tail of the hippocampus. The analysis with FPCA shows a clear disadvantage with that done with FICA because based on it we could determine the progress of the disease, whereas with FPCA there is not a progressive change in thresholding (color), i.e. it marks strongly where the differences are, but FICA gives greater details.

Tables 4 and 5 give the performance for the respective subproblem. Best or equal accuracies are achieved with the functional approach. Note that 100% correct classifications for CN vs AD are obtained by all methods except for SDA, but SVM uses 20 features, when the total number of subjects in that subproblem is 22.

It could be interesting to plot the scores of each subject on different components because these scatter plots can reveal interesting features, such as the distribution of the subjects on those components, clusters of subjects, outliers, etc. [4]. Note that the complex information in the hippocampi, which are structures in 3D, will be represented with simple scatter plots. For the MCI vs AD subproblem, which is the most difficult subproblem according to the obtained accuracies, we have computed the scores (the features used with LDA) using all the subjects in that subset (16) for the number of components selected ($J = 3$ for FPCA and $J = 4$ for FICA in this subproblem). In Figure 8 those scores for FPCA and FICA for the two components that visually best reflect the separation between groups are represented. The scatter plot for FICA shows a slight greater separation than that for FPCA, with two patients with different conditions nearly overlapped.

Instead of the scores used in the classification, in Figure 9 we show the discriminant values for the two discriminant functions for the three groups jointly, using all the subjects in that set (28) for the number of components selected ($J = 8$ for FPCA and $J = 9$ for FICA in this problem). Crosses, stars and circles represent the CN, MCI, and patients with early AD, respectively. The plots for FPCA and FICA are nearly identical. One of the MCI patient (the same in both plots) is near the AD group.

Table 6 shows the accuracies for each subproblem when the right hippocampi are analyzed (the results for the left hippocampi are also shown in order to make easier the comparisons). As in previous works [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16], it seems that the left hippocampi can discriminate better the AD condition than the right hippocampi.

References

[1] Hasboun, D., Chantôme, M., Zouaoui, A., Sahel, M., Deladoeuille, M., Sourour, N., Duyme, M., Baulac, M., Marsault, C., and Dormont, D. MR determination of hippocampal volume: Comparison of three methods. *American Journal of Neuroradioly*, 17:1091–1098, 1996.

[2] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning. Data mining, inference and prediction*. Springer-Verlag, second edition, 2009.

[3] Ramsay, J. O. and Silverman, B. W. *Applied Functional Data Analysis*. Springer, 2002.

[4] Jolliffe, I. T. *Principal Component Analysis*. Springer, second edition, 2002.

[5] Beg, M. F., Raamana, P. R., Barbieri, S., and Wang, L. Comparison of four shape features for detecting hippocampal shape changes in early Alzheimers. *Statistical Methods in Medical Research*, published online 30 May 2012(DOI: 10.1177/0962280212448975):1–24, 2012.

[6] Müller, M. J., Greverus, D., Weibrich, C., Dellani, P. R., Scheurich, A., Stoeter, P., and Fellgiebel, A. Diagnostic utility of hippocampal size and mean diffusivity in amnestic MCI. *Neurobiology of Aging*, 28(3):398–403, 2007.

[7] Epifanio, I. and Ventura-Campos, N. Functional data analysis in shape analysis. *Computational Statistics & Data Analysis*, 55(9):2758–2773, 2011.

[8] Wang, L., Beg, M. F., Ratnanather, J. T., Ceritoglu, C., Younes, L., Morris, J. C., Csernansky, J. G., and Miller, M. I. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE Trans. Med. Imaging*, 26(4):462–470, 2007.

[9] Eckerström, C., Olsson, E., Borga, M., Ekholm, S., Ribbelin, S., Rolstad, S., Starck, G., Edman, Å., Wallin, A., and Malmgren, H. Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia. The Göteborg MCI study. *Journal of the Neurological Sciences*, 272:48–49, 2008.

[10] Wolf, H., Grunwald, M., Kruggel, F., Riedel-Heller, S. G., Angerhöfer, S., Hojjatoleslami, A., Hensel, A., Arendt, T., and Gertz, H. Hippocampal volume discriminates between normal cognition; questionable and mild dementia in the elderly. *Neurobiol Aging*, 22(2):177–186, 2001.

[11] Woolard, A. and Heckers, S. Anatomical and functional correlates of human hippocampal volume asymmetry. *Psychiatry Res*, 201(1):48–53, 2012.

[12] Vijayakumar, A. and Vijayakumar, A. Comparison of hippocampal volume in dementia subtypes. *ISRN Radiology*, Article ID 174524(doi:10.5402/2013/174524):5, 2013.

[13] Carne, R. P., Vogrin, S., Litewka, L., and Cook, M. J. Cerebral cortex: an MRI-based study of volume and variance with age and sex. *J Clin Neurosci*, 13(1):60–72, 2006.

[14] Fox, N. C., Warrington, E. K., Freeborough, P. A., Hartikainen, P., Kennedy, A. M., Stevens, J. M., and Rossor, M. N. Presymptomatic hippocampal atrophy in Alzheimer's disease. a longitudinal MRI study. *Brain*, 119(6): 2001–7, 1996.

[15] Laakso, M. P., Partanen, K., Jr, Lehtovirta, M., Helkala, E. L., Hallikainen, M., Hanninen, T., Vainio, P., and Soininen, H. Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: An MRI study. *Neurology*, 46(3):678–681, 1996.

[16] Krasuski, J., Alexander, G., Horwitz, B., Daly, E., Murphy, D., Rapoport, S., and Schapiro, M. Volumes of medial temporal lobe structures in patients with Alzheimer's disease and mild cognitive impairment (and in healthy controls). *Biol Psychiatry*, 43(1):60–8, 1998.

[17] Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., and Colliot, O. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47 (4):1476–1486, 2009.

3. Figures and Tables

Table 1. Accuracies (%) for the CN vs MCI patients for the different values of L (the number of features between parentheses), selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the accuracy.

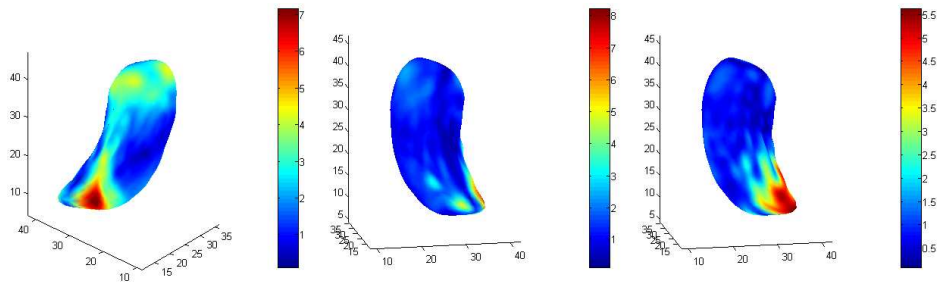| Method	$L=2$	$L=3$	$L=5$	$L=7$	$L=9$	$L=11$	$L=13$	$L=15$	$L=17$	$L=19$	$L=21$
FPCA	88.89 (4)	94.44 (4)	88.89 (4)	88.89 (4)	88.89 (4)	83.33 (4)	88.89 (5)	88.89 (5)	83.33 (4)	77.78 (4)	77.78 (4)
FICA	88.89 (5)	94.44 (5)	88.89 (4)	83.33 (4)	83.33 (4)	88.89 (6)	88.89 (6)	88.89 (6)	83.33 (5)	83.33 (5)	77.78 (4)

Figure 1. The effect of the first PC, the second PC and the third PC (from left to right) on the mean shape for 2 standard deviations of each component.
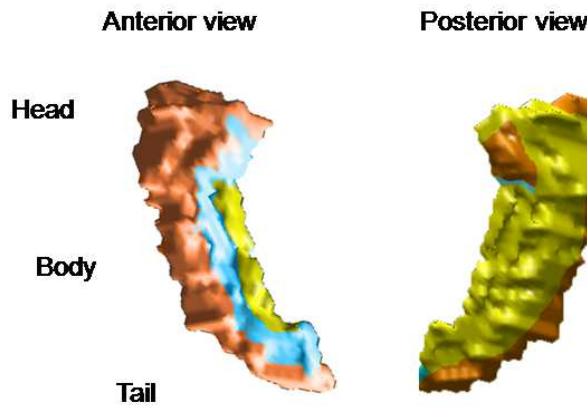

Figure 2. Zones on the left hippocampal surface. CA1 is in red, subiculum in yellow, and CA2, CA3, CA4 and gyrus dentatus in blue.

Table 2. Mean accuracies and their respective standard deviations between parentheses of the nested LOU, for the CN vs MCI patients, for different values of J.

Method	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	$J = 6$	$J = 7$	$J = 8$	$J = 9$	$J = 10$	$J = 11$	$J = 12$	$J = 13$
FPCA	72.88 (5.93)	71.24 (4.34)	82.35 (6.50)	91.18 (4.90)	92.48 (3.83)	86.93 (5.39)	83.66 (5.39)	83.99 (4.72)	80.39 (3.92)	76.14 (5.71)	74.18 (7.11)	73.86 (10.04)	74.51 (6.20)
FICA	60.78 (8.08)	69.28 (4.62)	72.55 (3.40)	87.91 (8.87)	87.58 (5.84)	88.56 (4.15)	85.95 (4.44)	80.07 (5.24)	78.76 (5.59)	73.53 (6.86)	67.65 (7.40)	66.01 (6.95)	66.34 (6.74)
SVM [17]	69.28 (8.45)	69.28 (8.67)	70.26 (9.09)	78.10 (6.74)	69.28 (10.84)	69.28 (9.71)	80.72 (6.73)	78.43 (8.77)	94.12 (3.40)	94.12 (3.40)	93.46 (3.33)	95.10 (2.94)	94.12 (3.40)
SDA	84.97 (4.48)	86.60 (6.14)	89.22 (6.58)	92.48 (6.45)	95.42 (5.02)	98.04 (3.40)	99.35 (2.69)	99.35 (2.69)	99.35 (2.69)	99.67 (1.35)	100 (0)	100 (0)	100 (0)

Table 3. Accuracies of the external LOU, for the CN vs MCI patients, for different values of J.

Method	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	$J = 6$	$J = 7$	$J = 8$	$J = 9$	$J = 10$	$J = 11$	$J = 12$	$J = 13$
FPCA	72.22	66.67	72.22	83.33	88.89	83.33	83.33	88.89	83.33	88.89	88.89	83.33	83.33
FICA	61.11	66.67	72.22	77.78	83.33	88.89	83.33	83.33	83.33	83.33	88.89	88.89	77.78
SVM [17]	61.11	61.11	61.11	61.11	50	50	50	50	61.11	61.11	50	77.78	61.11
SDA	72.22	72.22	72.22	72.22	72.22	72.22	72.22	72.22	72.22	72.22	72.22	77.78	77.78

Table 4. Performance for the MCI vs AD patients for the different methods, selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the accuracy.

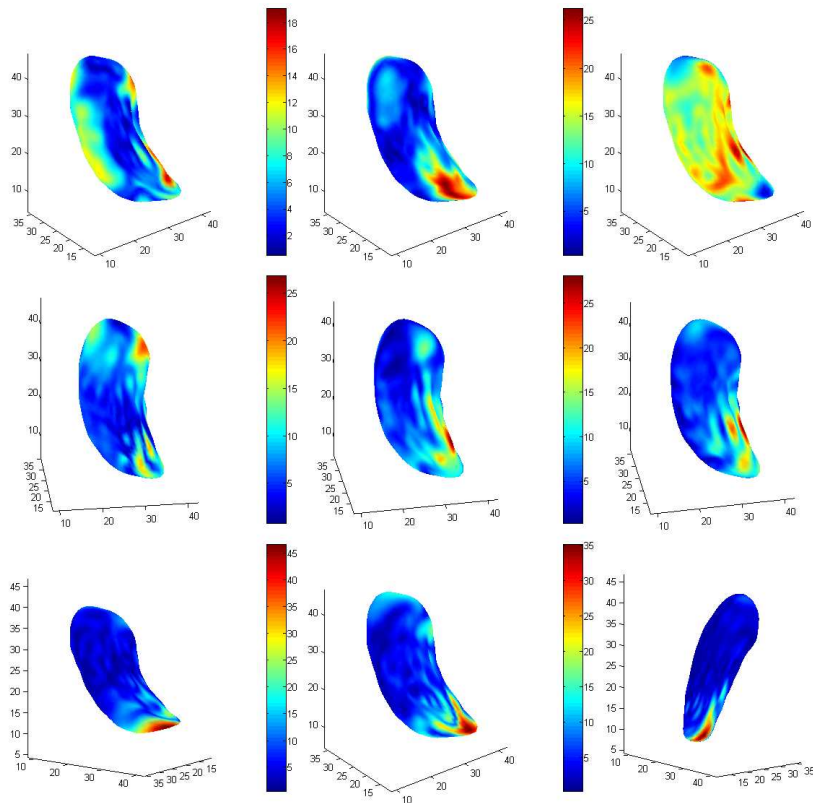| Method	Volume	FPCA	FICA	SVM [17]	SDA
Accuracy (%)	75.00	87.50	87.50	68.75	62.5
No. features	1	3	4	8	6
Sensitivity (%)	80.00	100	100	90.00	80.00
Specificity (%)	66.67	66.67	66.67	33.33	33.33

Figure 3. The effect of the first nine independent components on the mean shape.
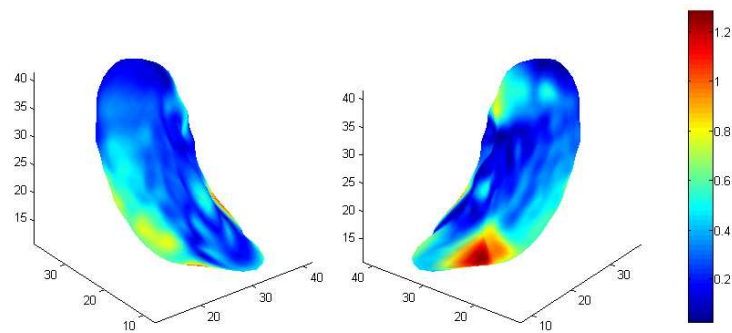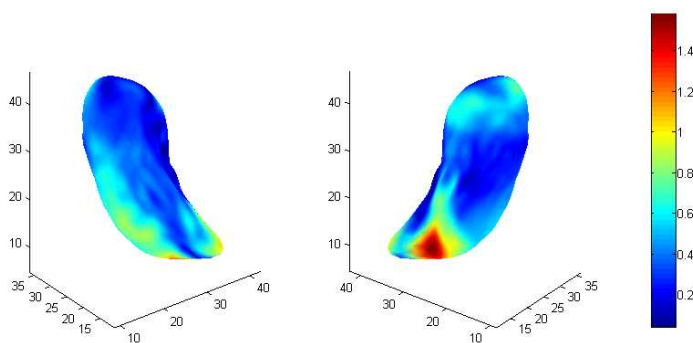

Figure 4. Functional linear discriminant with FICA for MCI vs AD ($J = 4$ components). The discriminant function significantly separates the groups (Wilks' $\Lambda = 0.320$, p-value $= 0.008$).

Table 5. Performance for the CN vs AD for the different methods, selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the accuracy.

Method	Volume	FPCA	FICA	SVM [17]	SDA
Accuracy (%)	100	100	100	100	90.91
No. features	1	1	2	20	6
Sensitivity (%)	100	100	100	100	90.00
Specificity (%)	100	100	100	100	91.67

Statist. Med. **0000**, 00 1–3
Prepared using *simauth.cls*

Copyright © 0000 John Wiley & Sons, Ltd.

www.sim.org   5

Figure 5. Functional linear discriminant with FICA for CN vs AD ($J = 2$ components). The discriminant function significantly separates the groups (Wilks' $\Lambda = 0.103$, p-value = 0).
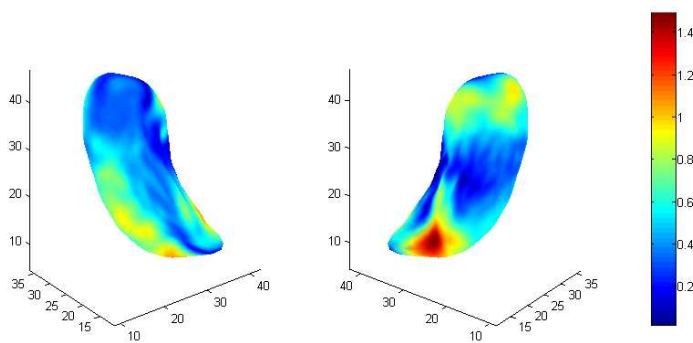

Figure 6. First functional linear discriminant with FICA for CN, MCI and AD ($J = 9$ components). The two discriminant functions significantly separate the groups at $\alpha = 0.05$, but only the first one, whose proportion of trace is 89.29%, is displayed (first: Wilks' $\Lambda = 0.046$, p-value = 0; second: Wilks' $\Lambda = 0.474$, p-value = 0.048).

Table 6. Accuracies (%) for the different subproblems for both hippocampi (right /left), selecting with double leave-one-out the features and using these features with leave-one-out for obtaining the accuracy. The number of features is between parentheses.

Method	3 groups	CN vs MCI	MCI vs AD	CN vs AD
FPCA	75 (1) / 85.71 (8)	72.22 (3) / 88.89 (5)	81.25 (5) / 87.5 (3)	100 (1) / 100 (1)
FICA	85.71 (4) / 85.71 (9)	66.67 (4) / 88.89 (6)	81.25 (4) / 87.5 (4)	95.45 (2) / 100 (2)

6 www.sim.org

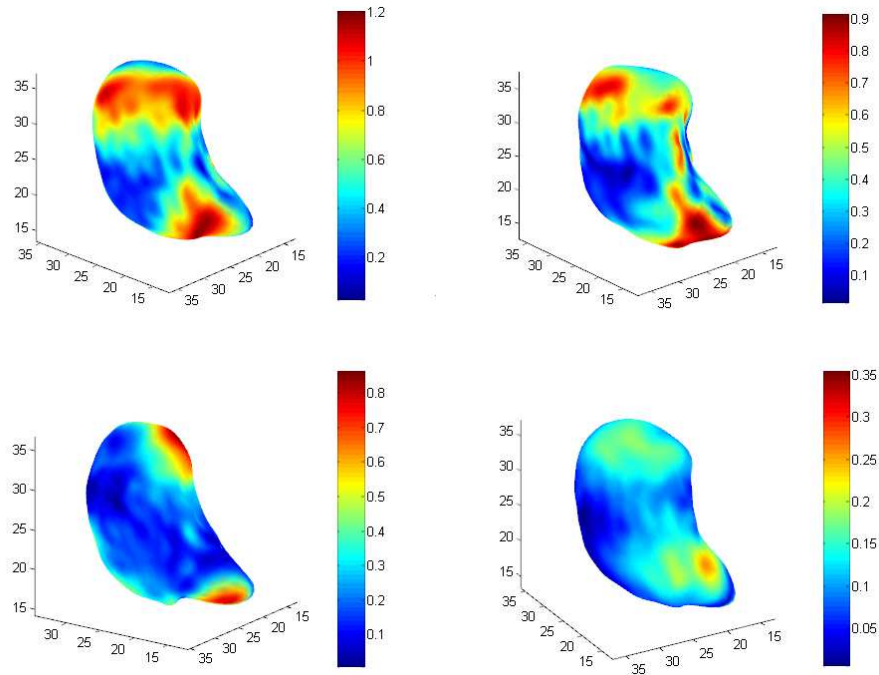*Statist. Med.* **0000**, 00 1–3

Prepared using **simauth.cls**

Figure 7. Top left: First functional linear discriminant with FPCA for CN, MCI and AD ($J = 8$ components), whose proportion of trace is 92.9% and Wilks' $\Lambda = 0.065$, p-value = 0 (note that for the second functional linear discriminant the p-value is higher than $\alpha = 0.05$, Wilks' $\Lambda = 0.609$, p-value = 0.154); Top right: Functional linear discriminant with FPCA for CN vs MCI ($J = 5$ components, Wilks' $\Lambda = 0.266$, p-value = 0.003); Bottom left: Functional linear discriminant with FPCA for MCI vs AD ($J = 3$ components, Wilks' $\Lambda = 0.612$, p-value = 0.105, note that the p-value is higher than $\alpha = 0.05$); Bottom right: Functional linear discriminant with FPCA for CN vs AD ($J = 1$ component, Wilks' $\Lambda = 0.959$, p-value = 0.367, note that the p-value is higher than $\alpha = 0.05$).
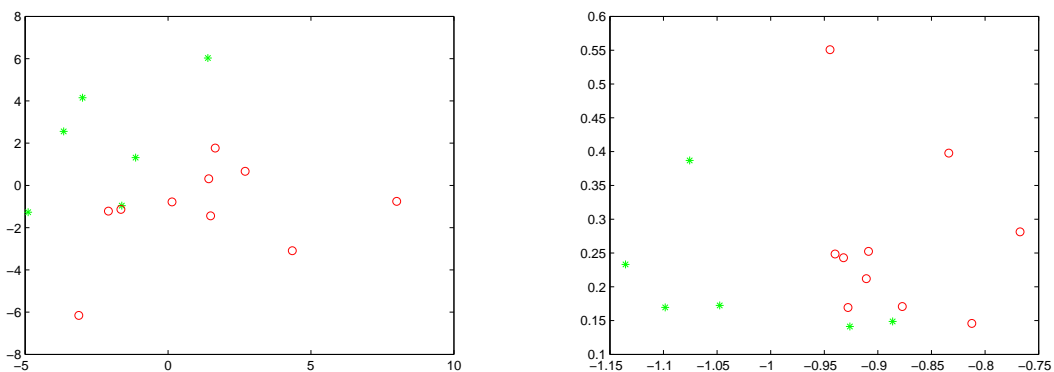

Figure 8. Scatter plot of scores for MCI vs AD. Component 1 vs 3 for FPCA (left) and component 2 vs 4 for FICA (right). Stars and circles represent the MCI and patients with early AD, respectively.
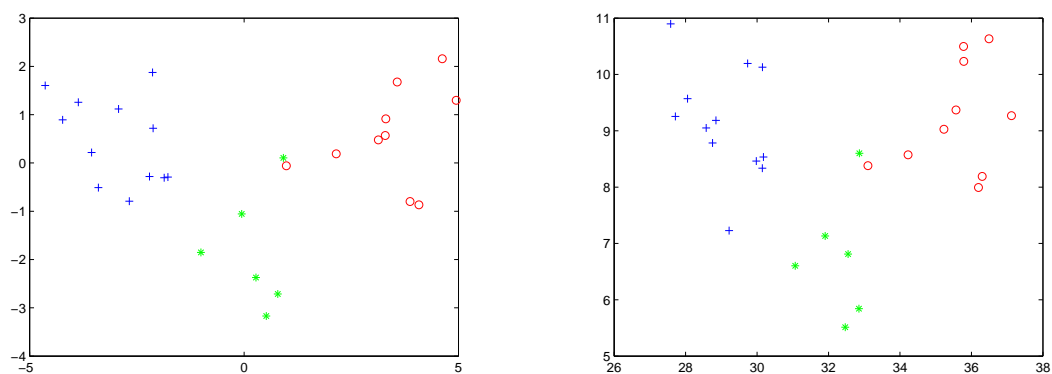
Figure 9. Scatter plot of discriminant values for the two discriminant functions of the three groups: FPCA (left) and FICA (right). Crosses, stars and circles represent the CN, MCI, and patients with early AD, respectively.